

Efficient Estimation of Resonant Coupling between Quantum Systems

Markku P.V. Stenberg,^{1,*} Yuval R. Sanders,² and Frank K. Wilhelm^{1,2}

¹*Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany*

²*IQC and Dept. of Physics and Astronomy, University of Waterloo,
200 University Ave. W, Waterloo, ON, N2L 3G1, Canada*

We present an efficient method for the characterization of two coupled discrete quantum systems, one of which can be controlled and measured. For two systems with transition frequencies ω_q , ω_r , and coupling strength g we show how to obtain estimates of g and ω_r whose error decreases exponentially in the number of measurement shots rather than as a power law expected in simple approaches. Our algorithm can thereby identify g and ω_r simultaneously with high precision in a few hundred measurement shots. This is achieved by adapting measurement settings upon data as it is collected. We also introduce a method to eliminate erroneous estimates with small overhead. Our algorithm is robust against the presence of relaxation and typical noise. Our results are applicable to many candidate technologies for quantum computation, in particular, for the characterization of spurious two-level systems in superconducting qubits or stripline resonators.

PACS numbers: 03.65.Wj, 03.65.Yz, 03.67.-a, 03.67.Lx

Parameter estimation in many microscopic and some macroscopic systems inevitably involves quantum measurements. This implies that parameters cannot be identified with a single measurement shot since the outcome of such a measurement is generally random. Instead, the standard approach is to determine ensemble averages for many experiments and fit the parameters of certain quantitative models to those averages. The most common example for this is spectroscopy: it involves direct measurement of the energy splittings between quantum states in the form of resonances to incoming radiation. Typically, a large ensemble average is produced by gathering data from a large number of independent trials, either simultaneously on an ensemble of molecules (in nuclear magnetic resonance [1]) or from many repetitions of a specific experiment (in optical spectroscopy of single molecules, quantum dots, or superconducting qubits [2]).

While being reliable in many contexts, this approach is often too resource intensive. Specifically, the error in the estimate of a single expectation value at a fixed measurement setting decreases in proportion to $M_r^{-\frac{1}{2}}$ after M_r measurement shots. Moreover, many choices of measurement settings are usually required for complex measurement tasks. Such slowness of parameter estimation can also turn into imprecision in the estimate if the parameters of interest drift as a function of time, broadening spectroscopic signatures.

Imprecision in system characterization is particularly problematic for quantum information processing applications. These require extremely precise logic operations, usually implemented as pulses. The pulse parameters such as length, amplitude, and carrier frequency, depend on the system parameters. In manufactured solid state qubits, this is rather central as they are subject to fabrication uncertainty.

In this Letter, we demonstrate that advanced spectroscopy can be performed far more efficiently. Our re-

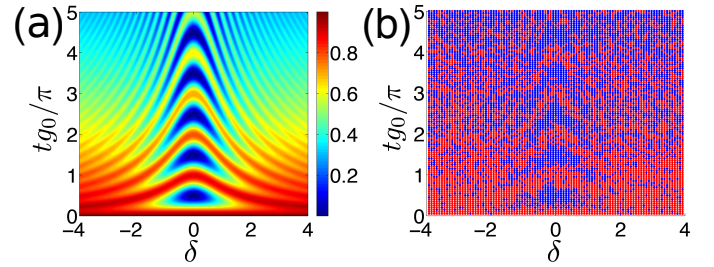


Figure 1: (color online) (a) Theoretically obtained swap spectrum in frequency-waiting time plane. Here, $\delta = (\omega_q - \omega_{r,0})/(2g_0)$, with ω_q the qubit frequency, $\omega_{r,0}$ the resonator frequency, and g_0 the coupling strength. The color scale of the swap spectrum represents the probability of the qubit being in its excited state. (b) Outcomes of a set of simulated single-shot measurements: blue the ground state; red the excited state.

sults are particularly relevant for the “tune-up” stage in quantum information processing, *i.e.*, for the initial calibration of the system and maintaining it. We use a controlled qubit to characterize the uncertain frequency of another mode ω_r that is coupled to the qubit with uncertain coupling strength g . Our algorithm employs modern Bayesian inference techniques to choose informative experimental settings while remaining computationally feasible. We demonstrate that our approach is robust against experimental imperfections [3, 4].

Our work belongs to the context of estimating an *a priori* unknown Hamiltonian [3–7]. Modern inference techniques have also been developed, *e.g.*, for phase estimation [8–17] and state estimation [18–20]. These techniques are able to use each bit of data obtained from experiment, instead of inferred expectation values only. This is both faster and more precise than the conventional approach involving ensemble averages.

Bayesian inference has been applied to identify qubit

Hamiltonians in a standard tomographic setup using fixed, evenly spaced waiting times t between preparation and measurement [21]. In extracting an unknown frequency of a qubit, significant advantage can be achieved through an adaptive algorithm [5, 6] that updates the measurement setting during the experimental data collection and tends to choose exponentially increasing, rather than evenly spaced, waiting times. Such a problem is equivalent to extracting g in a model discussed below when the frequency ω_r is known. The assumption that ω_r is known, however, limits the applicability of the model. The situation where both g and ω_r are initially unknown is more widely and practically applicable. To solve the latter problem, we deliver a strategy that chooses both t and the qubit frequency ω_q to achieve near-optimal scaling of errors in the estimates.

A qubit coupled to a resonator is described by the Jaynes-Cummings Hamiltonian

$$\hat{H}_{\text{JC}} = \frac{\hbar\omega_q}{2}\hat{\sigma}_z + \hbar\omega_r\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right) + \hbar g(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger), \quad (1)$$

where $g \ll \omega_r$. This model is broadly applicable to potential quantum computing technologies [22] and reproduces also the dynamics of spurious two-level systems [23–25], a notorious source of decoherence for superconducting qubits.

The standard method to estimate g and ω_r is called swap spectroscopy [26]. To understand swap spectroscopy, it helps to picture a single measurement setting as a single point in a figure such as Fig. 1(a) or 1(b). One starts by preparing the qubit in the excited state and the cavity in its ground state (typically by moving ω_q far away from ω_r and then exciting it with an external pulse). Then ω_q is fixed to a chosen value that in Fig. 1 determines the horizontal coordinate (see the caption). The system is allowed to evolve a time t (vertical coordinate of Fig. 1) after which the qubit is measured in the $\hat{\sigma}_z$ basis. The system is then reset to its ground state before the next measurement.

As the Jaynes-Cummings Hamiltonian conserves the total number of excitations $\hat{N} = \hat{a}^\dagger\hat{a} + \frac{1}{2}(\hat{\sigma}_z + \hat{1})$ we can describe the excitation oscillating between the qubit and the resonator in the single-excitation subspace by the Hamiltonian $\hat{H}' = \frac{\Delta\omega}{2}\hat{\eta}_z + g\hat{\eta}_x$, where $\hat{\eta}$ are Pauli matrices and the detuning $\Delta\omega = \omega_q - \omega_r$. This form of the relevant Hamiltonian occurs, in addition to the applications mentioned above, also in magnetic resonance spectroscopy. For the Hamiltonian \hat{H}' the probability of the qubit being in its excited state is

$$P_{\omega_q,t}(1|g,\omega_r) = \frac{1}{2}\left(\frac{4g^2}{\omega_R^2}\cos\omega_R t + 1 + \frac{\Delta\omega^2}{\omega_R^2}\right), \quad (2)$$

with $\omega_R = \sqrt{\Delta\omega^2 + 4g^2}$. The Supplemental Material generalizes this formula to account for qubit relaxation.

In conventional swap spectroscopy, the measurement is repeated at a setting (ω_q, t) in order to establish an

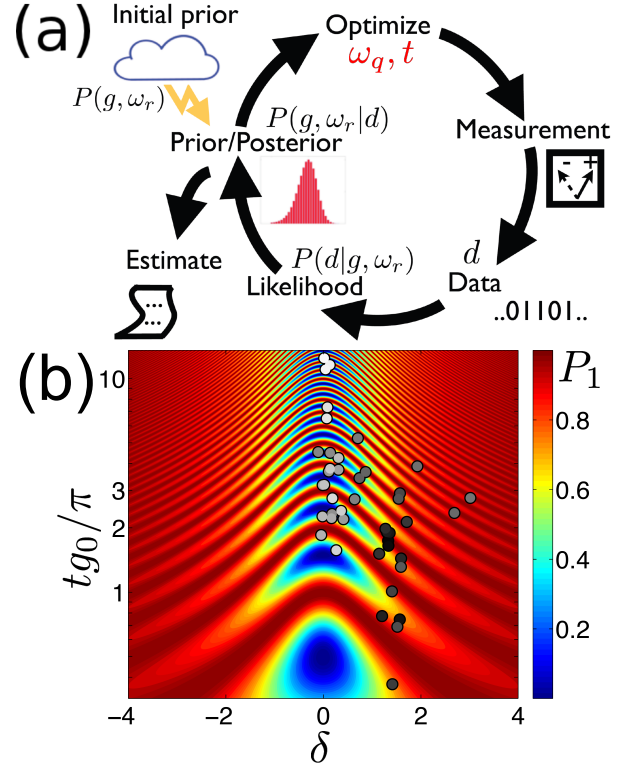


Figure 2: (color online) (a) Illustration of an adaptive Bayesian inference scheme. (b) Exemplary trace of adaptive measurements overlaid on swap spectrum in the frequency-waiting time plane. The order of the measurement shots is denoted by their color (from black to white). Here $\delta = (\omega_q - \omega_{r,0})/(2g_0)$, with ω_q the qubit frequency, $\omega_{r,0}$ the resonator frequency, and g_0 the coupling strength. The color scale for the swap spectrum represents the probability for the qubit to be in its excited state. Note the logarithmic scale on the time axis.

ensemble to calculate the relative frequency of the excited state and, hence, approximate its probability. To estimate g , usual swap spectroscopy first seeks ω_q , where $\Delta\omega = 0$, and where the Chevron pattern of the oscillating excitation probability has a maximum visibility; cf. Fig. 1. Measuring the angular frequency of these oscillations then yields $2g$ by Eq. (2).

Our algorithm, rather than establishing ensemble averages, chooses a new pair (ω_q, t) after each measurement step. We employ Bayes' theorem, which reads

$$P(g, \omega_r | d) = \frac{P(d | g, \omega_r) P(g, \omega_r)}{P(d)}, \quad d = 0, 1 \quad (3)$$

for our system. This formula can be understood as a rule for iterative learning of the parameters. One starts with an initial probability distribution, or prior, $P(g, \omega_r)$ that describes one's *a priori* conception about the uncertain parameters. Based on the measurement outcome d , cf. Fig. 2(a), one can use Bayes' theorem to update the probability distribution into a posterior based on the

likelihood of the data $P(d|g, \omega_r)$ according to conjectured model parameters g and ω_r . The normalization factor in the denominator can be calculated via the integral $P(d) = \int P(d|g, \omega_r)P(g, \omega_r)dg d\omega_r$. The estimate is obtained from the mean value of the posterior. The posterior is then identified as the prior for the next measurement. Thus each measurement outcome is immediately incorporated into our knowledge of the system.

We now want to optimize the measurement settings based on the current knowledge about ω_r and g . In principle, for each adaptation step one can maximize utility (e.g., a negative trace of the posterior covariance matrix or information gain I [27]) of the next shot as if it were the last in the series, a so-called “greedy” algorithm. Computationally however, maximizing utility between the shots is quite unwieldy. Moreover, to optimize the whole series of measurement shots, *i.e.*, to optimize globally, it is not sufficient to optimize greedily, *i.e.*, locally. We have therefore studied information gain $I(\omega_q, t)$ as a function of the control parameters with different parameter values (g, ω_r) and different priors. These considerations (which we carried out off-line rather than in real time parallel with parameter estimation) suggest a strategy in which $t \sim 1/\sigma_g$ and $\omega_q - \mu_\omega \sim \sigma_\omega$. With respect to the prior, μ_ω is the mean of ω_r and σ_g (σ_ω) is the standard deviation of g (ω_r). Our measurement strategy chooses the M th measurement setting (ω_q, t) according to the rule

$$t = \begin{cases} \frac{ar_1}{\sigma_g} & \text{if } M \leq M_0 \\ \frac{|a+bz|}{\sigma_g} & \text{if } M > M_0 \end{cases},$$

$$\omega_q = \begin{cases} \mu_\omega + (r_2 - \frac{1}{2})\mu_g & \text{if } M \leq M_0 \\ \mu_\omega + c(r_2 - \frac{1}{2})\sigma_\omega & \text{if } M > M_0. \end{cases} \quad (4)$$

Here, $a = 1.57$, $b = 0.518$, and $c = 3.0$ are numerical constants that we have found to yield a robust and efficient strategy. Furthermore, z is a standard normal deviate and $r_{1,2}$ are uniform random variables on the interval $[0, 1]$. This strategy is the central result of our Letter. The measurement settings with $M \leq M_0$ are chosen more uniformly to obtain a unimodal posterior in the beginning of the series [6], which makes the strategy more effective. We choose $M_0 = 15$ which we found to be sufficient especially for our prior and the parameters considered in Figs. 3, 4(a), and 4(b). Figure 2(b) illustrates an exemplary trace of measurement shots chosen adaptively according to Eq. (4). Even though the algorithm adaptively chooses measurements based on the current uncertainty in the Hamiltonian, it is different from “particle guess heuristic” (PGH) of Refs. [4, 7] since it does not attempt to counteract the time evolution $e^{-i\hat{H}_J c t}$ or the influence of g . In particular, our algorithm only controls (ω_q, t) . These control knobs are easier to adjust experimentally than the 3 degrees of freedom including two-qubit gates suggested by PGH.

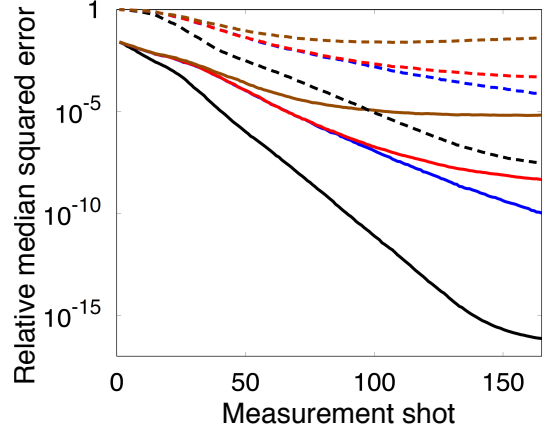


Figure 3: (color online) Relative median squared error of the estimate of g (solid) and ω_r (dashed) calculated from an ensemble of 10 000 simulated samples (see text). The curves correspond to $T_1 = \infty$ and $P_e = 0$ (black), $T_1 = \infty$ and $P_e = 0.1$ (blue), $T_1\mu_g = 2000\pi$ and $P_e = 0.1$ (red), and $T_1\mu_g = 40\pi$ and $P_e = 0.1$ (brown). Here, T_1 is the relaxation time, P_e is the probability for a readout error, and μ_g the mean of g over the ensemble.

A numerical challenge with Bayesian inference techniques is that each application of Bayes’ theorem (3) requires the evaluation of the computationally expensive integral $P(d)$. To calculate the integrals we adapt a sequential Monte Carlo approach [3, 18, 28–30] with moving grid points or “particles” whose density describes the probability distribution of interest [31]. In evaluating a probability distribution in fixed grid points, the density of the grid points would limit the precision on the estimates, which is mitigated by the adaptive grid. We perform the calculations with 50 000 particles, so, for instance, in computing $P(d)$ one evaluates 50 000 times the likelihood function.

To consider the performance of our algorithm, we have applied it to ensembles of 10 000 simulated samples with randomly chosen parameters $(g_0, \omega_{r,0})$. Here the subscript 0 denotes a specific fixed true value, in contrast to the symbols naming a quantity. The values g_0 have been chosen from a log-normal distribution with the mean μ_g and the standard deviation σ_g while $\omega_{r,0}$ have been chosen from the normal distribution $\mathcal{N}(\mu_\omega, \sigma_\omega)$. For each sample we have chosen the initial prior of the (g, ω_r) estimate to coincide with the probability distribution from which the true values $(g_0, \omega_{r,0})$ are randomly picked. Unless specified otherwise we have chosen $\sigma_g = 0.25\mu_g$ and $\sigma_\omega = \mu_g$.

Figure 3 exhibits the relative median squared error as a function of measurement shots for g (solid) and ω_r (dotted). The changes in parameters represent the effects of relaxation and noise: (i) $T_1 = \infty$, $P_e = 0$ (black), (ii) $T_1 = \infty$, $P_e = 0.1$ (blue), (iii) $T_1 = \frac{2000\pi}{\mu_g}$, $P_e = 0.1$ (red), and (iv) $T_1 = \frac{40\pi}{\mu_g}$, $P_e = 0.1$ (brown). Here T_1 is

Table I: Number of outliers per 10 000 simulated samples with $T_1 = \infty$, $P_e = 0$. Rows correspond to the number of outliers with squared error larger than $\tilde{\mathcal{E}}_g^2$ after a given number of measurement shots (indicated by the columns).

$\tilde{\mathcal{E}}_g^2/\text{Shots}$	150	300	600	1200	3600	6600
10^{-10}	533	466	276	25	3	0
10^{-7}	265	251	111	18	2	0
10^{-4}	118	116	25	14	1	0

the relaxation time and P_e the probability of a readout error. The error decreases exponentially with the base of the exponential function affected by P_e . When the error reaches a low crossover value determined by T_1 , the decay law crosses over into a power law.

For each simulated ensemble there are some samples we call “outliers” for which the error significantly exceeds the exponentially decaying median and the width of the posterior distribution. The outliers mostly correspond to unlikely parameter values. Our algorithm can be made robust against such outliers through repetition as follows, based on the idea that most bad estimates are good initial guesses. After 300 measurement shots we set the prior widths back to their original values but keep the mean of the probability distribution unchanged. Another 300 measurement shots are performed thereafter. We then compare the estimates after 300 and 600 measurement shots. If their difference is smaller than a set threshold, we conclude that we have found a correct estimate, otherwise we start a new search of the estimate. For the new search we choose a prior whose mean values for g and ω_r are randomly chosen from the original prior while the prior widths equal those of the original prior. Table I summarizes the performance of our outlier correction scheme. Outliers are defined as the samples with the squared error of the g estimate larger than threshold $\tilde{\mathcal{E}}_g^2$. Our scheme appears to reduce the number of outliers with an acceptable overhead.

Figure 4 exhibits the average number of measurement shots required to meet the desired level of relative mean squared error. Each point represents the average performance over 10 000 samples. The robustness of our algorithm against measurement errors and relaxation is demonstrated in Figs. 4(a) and 4(b), and the effect of greater initial parameter uncertainty is considered in Figs. 4(c) and 4(d). We find that when $|\omega_{r,0} - \mu_\omega|$ is no larger than few times g_0 , the number of measurement shots is comparable to the one ideally required when ω_r is initially known precisely [3, 6], indicating the near optimality of our algorithm in this parameter region.

With conventional swap spectroscopy, the choice of the measurement settings tends to limit the accuracy of the estimates. Their squared error scales no better than $\sim M_s^{-1}$ in the number of measurement settings M_s (grid points distributed on a rectangular grid in Fig. 1) and in

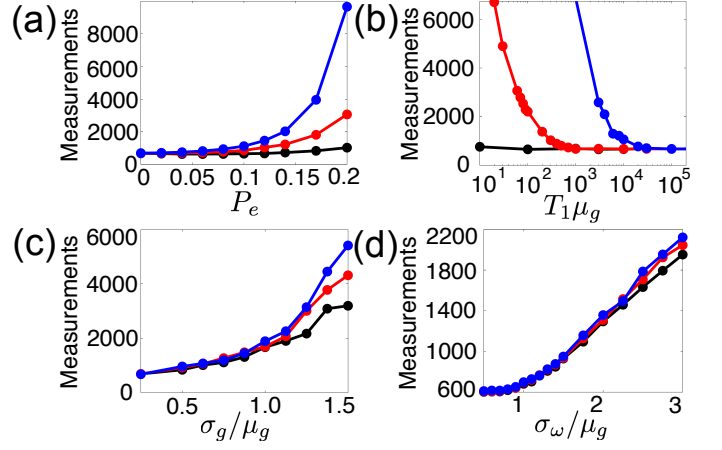


Figure 4: (color online) The average number of measurement shots required to meet the desired level of relative mean squared error \mathcal{E}_g^2 : 10^{-4} (black), 10^{-7} (red), and 10^{-10} (blue). The number of measurements presented as a function of (a) probability of readout error P_e , (b) relaxation time T_1 , (c) standard deviation σ_g over the initial prior, and (d) standard deviation σ_ω over the initial prior.

a typical experiment M_s is in the range $O(10^4) - O(10^5)$. The accuracy of the estimate of $P_{t,\omega_q}(1|g, \omega_r)$ [color scale in Fig. 1(a)] is also limited due to projection noise as the ensemble average is typically taken over $M_r \sim O(10^3) - O(10^4)$ trials. Our results show that it is possible to obtain far more accurate estimates with a much smaller number of measurement shots which makes our approach highly efficient.

In order to estimate the experimental time advantage, one has to take into account the delay of requesting and transferring data between the experiment and the computation that determines the measurement setting t_{latency} , as well as the time to computationally determine a single measurement setting t_{adapt} . If the adaptive scheme is implemented on a field-programmable gate array, $t_{\text{latency}} \sim 200 \mu\text{s}$ is dominated by $t_{\text{adapt}} \sim 10 \text{ ms}$ [32]. The typical time to initialize the system and perform the measurement, e.g., in superconducting qubits is $t_{\text{cycle}} \sim 100 \mu\text{s}$. We therefore expect it is optimal to repeat measurements in a single setting $\sim 10 - 100$ times. Making several repetitions decreases the number of required adaptation steps. We estimate that in a few seconds it is possible to obtain estimates that are more precise by several orders of magnitude compared to conventional swap spectroscopy where the time to carry out $M_s M_r$ measurement shots ranges from minutes to several hours. For experimental considerations it will also be useful to develop a more general algorithm to choose measurement settings for wide frequency ranges containing multiple resonators at frequencies $\omega_r^{(i)}$.

In conclusion, by adaptively focusing measurements on the regions of high information gain we obtained a globally efficient measurement strategy for two control pa-

rameters. Our algorithm makes advanced spectroscopy drastically more efficient.

We thank D. Sank, J. Kelly, and J.M. Martinis for numerous discussions. F.W. acknowledges early discussions with C. Bender, H. Zähle, and M. Hein. Work supported by the European Union through ScaleQIT as well as by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U. S. Government.

* markku.stenberg@iki.fi

- [1] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance* (Wiley, Chichester, 2008), 2nd ed.
- [2] W. Demtröder, *Laser Spectroscopy*, Vol. 1 (Springer, Berlin, 2008), 4th ed.
- [3] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, *New J. Phys.* **14**, 103013 (2012).
- [4] N. Wiebe, C. E. Granade, C. Ferrie, and D. G. Cory, *Phys. Rev. A* **89**, 042314 (2014).
- [5] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, *Phys. Rev. A* **84**, 052315 (2011).
- [6] C. Ferrie, C. E. Granade, and D. G. Cory, *Quant. Inf. Proc.* **12**, 611 (2013).
- [7] N. Wiebe, C. E. Granade, C. Ferrie, and D. G. Cory, *Phys. Rev. Lett.* **112**, 190501 (2014).
- [8] V. Giovannetti, S. Lloyd, and L. Maccone, *Science* **306**, 1330 (2004).
- [9] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, *Nature (London)* **450**, 393 (2007).
- [10] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, *Phys. Rev. A* **80**, 052114 (2009).
- [11] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **104**, 063603 (2010).
- [12] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde, *Nat. Photon.* **5**, 43 (2011).
- [13] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **107**, 233601 (2011).
- [14] J. Joo, W. J. Munro, and T. P. Spiller, *Phys. Rev. Lett.* **107**, 083601 (2011).
- [15] H. Yonezawa *et al.*, *Science* **337**, 1514 (2012).
- [16] G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, *Nat. Nanotechnol.* **7**, 105 (2012).
- [17] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, *Nat. Nanotechnol.* **7**, 109 (2012).
- [18] F. Huszár and N. M. T. Houlshby, *Phys. Rev. A* **85**, 052120 (2012).
- [19] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, *Phys. Rev. Lett.* **109**, 130404 (2012).
- [20] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlshby, F. Huszár, and S. P. Kulik, *Phys. Rev. A* **87**, 062122 (2013).
- [21] S. G. Schirmer and D. K. L. Oi, *Phys. Rev. A* **80**, 022333 (2009).
- [22] S. Haroche and J.-M. Raimond, *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford University Press, New York, 2006); L. S. Bishop, J. M. Chow, J. Koch, A. A. Houck, M. H. Devoret, E. Thuneberg, S. M. Girvin, and R. J. Schoelkopf, *Nat. Phys.* **5**, 105 (2009); A. Imamoglu, D. D. Awschalom, G. Burkard, D. P. DiVincenzo, D. Loss, M. Sherwin, and A. Small, *Phys. Rev. Lett.* **83**, 4204 (1999).
- [23] M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, N. Katz, E. Lucero, A. O’Connell, H. Wang, A. N. Cleland, and J. M. Martinis, *Nat. Phys.* **4**, 523 (2008).
- [24] T. Palomäki *et al.*, *Phys. Rev. B* **81**, 144503 (2010).
- [25] Y. Shalibo, Y. Roife, D. Shwa, F. Zeides, M. Neeley, J. M. Martinis, and N. Katz, *Phys. Rev. Lett.* **105**, 177001 (2010).
- [26] M. Mariani *et al.*, *Science* **334**, 61 (2011).
- [27] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, New York, 2006).
- [28] M. West, *J. Roy. Stat. Soc. B Met.* **55**, 409 (1993).
- [29] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, *Radar and Signal Processing IEE Proc.-F* **140**, 107 (1993).
- [30] J. Liu and M. West in, *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. Freitas, and N. Gordon (Springer, New York, 2001).
- [31] We modify the algorithm, e.g., by *resampling* particles from log-normal rather than normal (cf. Eq. (3.6) in [30]) distribution to avoid negative values of g .
- [32] D. Sank, J. Kelly, and J. Martinis (private communication).

Supplemental Material

The dynamics of the qubit-resonator system is governed by the Master equation

$$\dot{\hat{\rho}}_S = -i [\hat{H}, \hat{\rho}_S] + \mathcal{D} [\hat{A}] \hat{\rho}_S, \quad (5)$$

with $\hat{\rho}_S$ the density matrix in the Schrödinger picture. In the absence of relaxation the system is described by the Jaynes-Cummings Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_1, \quad \hat{H}_0 = \frac{E_q}{2} \hat{\sigma}_z + \frac{E_r}{2} \hat{\tau}_z, \quad \hat{H}_1 = g (\hat{\sigma}_+ \hat{\tau}_- + \hat{\sigma}_- \hat{\tau}_+). \quad (6)$$

Here E_q and E_r , are the bare qubit and resonator energies, respectively, and the qubit-resonator coupling is characterized by g . Furthermore, $\hat{\sigma}_z$ ($\hat{\tau}_z$) are Pauli matrices in qubit (resonator) subspace, whereas $\hat{\sigma}_\pm$ and $\hat{\tau}_\pm$ are raising and lowering operators. We focus on energy relaxation of the qubit as the main channel of loss, described by a Lindblad form

$$\mathcal{D} [\hat{A}] \hat{\rho}_S \equiv \hat{A} \hat{\rho}_S \hat{A}^\dagger - \frac{1}{2} \{ \hat{A}^\dagger \hat{A}, \hat{\rho}_S \} \quad (7)$$

with the amplitude damping Lindblad operator

$$\hat{A} = \sqrt{\Gamma} \hat{\sigma}_- \quad (8)$$

where Γ is the relaxation rate.

We would first like to show that we can drop the state $|11\rangle$ in the equations from the very beginning. We switch to the interaction picture with the Schwinger-Tomonaga equation

$$\dot{\hat{\rho}}_I = -i [\hat{H}_1, \hat{\rho}_I] + \mathcal{D} [\hat{A}_I] \hat{\rho}_I. \quad (9)$$

It is easy to show that $\hat{H}_1|11\rangle=0$. Moreover, $\hat{A}_I = e^{-iE_q t} \hat{A}$ implies $\hat{A}_I \hat{\rho}_I \hat{A}_I^\dagger |11\rangle = 0$. Thus all the time-dependence left is

$$\left\langle 11 \left| \dot{\hat{\rho}}_I \right| 11 \right\rangle = -\Gamma \langle 11 | \hat{\rho}_I | 11 \rangle. \quad (10)$$

Therefore if the initial state does not include a $|11\rangle$ component, that state stays unoccupied.

Below we assume that the initial state at $t = 0$ is $|10\rangle$. As the Jaynes-Cummings Hamiltonian preserves the number of excitations we can describe the resulting dynamics in the single-excitation subspace by the Hamiltonian $\hat{H}' = \frac{\Delta\omega}{2} \hat{\eta}_z + g \hat{\eta}_x$. Here $\hat{\eta}$ are Pauli matrices and $\Delta\omega = \omega_q - \omega_r$ is the detuning frequency. The eigenstates of \hat{H}' are

$$\begin{aligned} |e\rangle &= \cos \frac{\theta}{2} |10\rangle + \sin \frac{\theta}{2} |01\rangle, \\ |g\rangle &= -\sin \frac{\theta}{2} |10\rangle + \cos \frac{\theta}{2} |01\rangle, \end{aligned} \quad (11)$$

where $\tan \theta = \frac{2g}{\Delta\omega}$. The eigenenergies are $E = \pm \frac{\hbar\omega_R}{2}$, with $\omega_R = \sqrt{\omega^2 + 4g^2}$. Note that $\Delta\omega = 0$ corresponds to $\theta = \pi/2$.

To treat relaxation, we now express the qubit lowering operator in these eigenstates as

$$\begin{aligned} \hat{\sigma}_- &= |0\rangle\langle 1| \otimes \hat{1} \\ &= |00\rangle\langle 01| \left(\cos \frac{\theta}{2} |e\rangle\langle e| - \sin \frac{\theta}{2} |g\rangle\langle g| \right) + \\ &\quad + \left(\sin \frac{\theta}{2} |e\rangle\langle g| + \cos \frac{\theta}{2} |g\rangle\langle e| \right) |11\rangle\langle 11|. \end{aligned} \quad (12)$$

Our approach is to perform a rotating wave approximation (RWA), recognizing that $\omega_R/2 \geq g \gg \Gamma$. To carry out the transformation, we first introduce a further interaction picture where the perturbation now includes only the relaxation

$$\dot{\hat{\rho}}_{I'} = \mathcal{D} [\hat{A}_{I'}] \hat{\rho}_{I'}. \quad (13)$$

In this picture the lowering operator in the qubit subspace is

$$\begin{aligned} \hat{\sigma}_-^{(I')} &= e^{-i(E_q+E_r)t/2\hbar} |00\rangle\langle 00| \left(\cos \frac{\theta}{2} e^{-i\omega_R t/2} |e\rangle\langle e| \right. \\ &\quad \left. - \sin \frac{\theta}{2} e^{i\omega_R t/2} |g\rangle\langle g| \right) + e^{-i(E_q+E_r)t/2\hbar} \\ &\quad \times \left(\sin \frac{\theta}{2} e^{i\omega_R t/2} |e\rangle\langle g| + \cos \frac{\theta}{2} e^{-i\omega_R t/2} |g\rangle\langle e| \right) |11\rangle\langle 11|. \end{aligned} \quad (14)$$

To evaluate the Lindblad form on the right hand side of Eq. (13) we need to transform the operator $\hat{\sigma}_+ \hat{\sigma}_- = |1\rangle\langle 1| \otimes \hat{1}$ to the interaction picture. In the RWA the transformed operator reads

$$\left(\hat{\sigma}_+^{(I')} \hat{\sigma}_-^{(I')} \right)_{\text{RWA}} = \cos^2 \frac{\theta}{2} |e\rangle\langle e| + \sin^2 \frac{\theta}{2} |g\rangle\langle g| + |11\rangle\langle 11|. \quad (15)$$

We can now drop all the terms involving the state $|11\rangle$ using the same argument as in the context of Eq. (10). Hence we obtain

$$\begin{aligned} \left\langle e \left| \dot{\hat{\rho}}_{I'} \right| e \right\rangle &= -\cos^2 \frac{\theta}{2} \Gamma \langle e | \hat{\rho}_{I'} | e \rangle, \\ \left\langle g \left| \dot{\hat{\rho}}_{I'} \right| g \right\rangle &= -\sin^2 \frac{\theta}{2} \Gamma \langle g | \hat{\rho}_{I'} | g \rangle, \\ \left\langle e \left| \dot{\hat{\rho}}_{I'} \right| g \right\rangle &= -\frac{\Gamma}{2} \langle e | \hat{\rho}_{I'} | g \rangle. \end{aligned} \quad (16)$$

We would now like to calculate the evolution of the population of $|10\rangle$ starting in that state. The initial density matrix is

$$\begin{aligned} \hat{\rho}_{I'}(t=0) &= |10\rangle\langle 10| \\ &= \cos^2 \frac{\theta}{2} |e\rangle\langle e| + \sin^2 \frac{\theta}{2} |g\rangle\langle g| - \frac{1}{2} \sin \theta (|e\rangle\langle g| + |g\rangle\langle e|). \end{aligned} \quad (17)$$

Using Eqs. (13) and (16) we find that the state decays according to

$$\begin{aligned} \hat{\rho}^I(t) &= \cos^2 \frac{\theta}{2} e^{-\Gamma t \cos^2 \theta/2} |e\rangle\langle e| + \sin^2 \frac{\theta}{2} e^{-\Gamma t \sin^2 \theta/2} |g\rangle\langle g| \\ &\quad - \frac{1}{2} \sin \theta e^{-\Gamma t/2} (|e\rangle\langle g| + |g\rangle\langle e|). \end{aligned} \quad (18)$$

The projection operator $\hat{\Pi} = |10\rangle\langle 10|$ corresponding to the measurement is in the interaction picture

$$\begin{aligned} \hat{\Pi}_{I'}(t) = & \cos^2 \frac{\theta}{2} |e\rangle\langle e| + \sin^2 \frac{\theta}{2} |g\rangle\langle g| \\ & - \frac{1}{2} \sin \theta (|e\rangle\langle g| e^{i\omega_R t} + |g\rangle\langle e| e^{-i\omega_R t}). \end{aligned} \quad (19)$$

Since the trace of an operator is the same in the interaction and the Schrödinger pictures we find the occupation probability of the state $|10\rangle$

$$\begin{aligned} P_{t,\omega_q}(1|g, \omega_r) = & \text{Tr}(\hat{\Pi}\hat{\rho}) = \frac{(1 + \cos \theta)^2}{4} e^{-\Gamma t(1 + \cos \theta)/2} \\ & + \frac{(1 - \cos \theta)^2}{4} e^{-\Gamma t(1 - \cos \theta)/2} + \frac{\sin^2 \theta}{2} e^{-\Gamma t/2} \cos \omega_R t. \end{aligned} \quad (20)$$

Using $\tan \theta = \frac{2g}{\Delta\omega}$ and $T_1 = \frac{1}{\Gamma}$ this can be recast as

$$\begin{aligned} P_{t,\omega_q}(1|g, \omega_r) = & \left(\frac{\omega_R + \Delta\omega}{2\omega_R} \right)^2 e^{-(\omega_R + \Delta\omega)t/2\omega_R T_1} \\ & + \left(\frac{\omega_R - \Delta\omega}{2\omega_R} \right)^2 e^{-(\omega_R - \Delta\omega)t/2\omega_R T_1} + \frac{2g^2}{\omega_R^2} e^{-t/2T_1} \cos \omega_R t. \end{aligned} \quad (21)$$

This generalizes Eq. (2) in the main text for finite relaxation.